AQUA
Water Infrastructure, Ecosystems and Society

AQUA — Water Infrastructure, Ecosystems and Society Vol 72 No 12, 2422 doi: 10.2166/aqua.2023.317

A study on the consumption of safe drinking water and its determinants using the weighted multinomial logistic model

Maher Ali*, Said Farooq Shah, Yasir Khan, Syed Muhammad Asim and Syed Fasih Ullah
Department of Statistics, University of Peshawar, Peshawar, Pakistan
*Corresponding author. E-mail: maherali80005@gmail.com

ABSTRACT

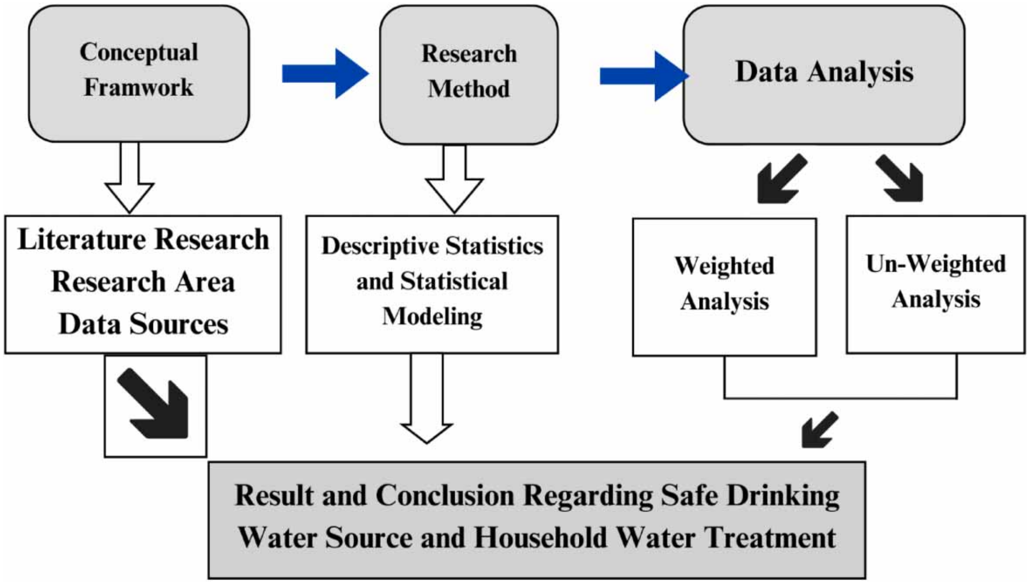
This study introduces a novel approach by incorporating weighted analyses along with unweighted to comprehensively examine the decision-making processes related to drinking water sources and treatment in Pakistani households, utilizing 2018 Pakistan Demographic and Health Survey (PDHS) data for weighted descriptive statistics and weighted multinomial logistic models. This research uncovers distinct preferences between urban and rural regions, with urban areas favoring piped water and rural areas relying on surface water and unprotected wells. The age of the household head, household size, media exposure, education, women's empowerment, and wealth are identified as crucial factors influencing these choices. Furthermore, this study delves into household water treatment (HWT) adoption and finds varying rates among rural and urban areas, with household head age and education level playing pivotal roles. Remarkably, household size and media exposure exhibit no significant impact on HWT adoption. The incorporation of weighted and unweighted analyses enriches the understanding of safe drinking water preferences, highlighting regional disparities and influential factors in water source and treatment method adoption within Pakistani households.

Key words: bottled/filter water, drinking water sources, household water treatment, protected well, safe drinking water, surface water

HIGHLIGHTS

- Use of sampling weights.
- Checking the impact of sampling weights on determinants of safe drinking water.
- Finding out the determinants of household water treatment.
- Computing the descriptive statistics of safe drinking water and household water treatment (HWT).
- Comparing the weighted and unweighted descriptive statistics based on their properties.

GRAPHICAL ABSTRACT



1. INTRODUCTION

It is a fundamental human right to have access to clean drinking water, and providing this basic human need is a crucial first step in raising living standards (Dinka 2018). One of the most important factors in eliminating poverty is access to safe drinking water, a basic health requirement (Zulfiqar *et al.* 2016). Along with increasing diseases and mortality, inadequate drinking water also raises healthcare costs, decreases productivity at work, and ultimately pushes people into poverty (Ahmad & Sattar 2010). Water is the basic element of the environment supporting life on Earth. Although more than 70% of the Earth's surface is covered with water, less than 1% of it may be utilized as pure water for drinking. On our 'blue globe,' the ocean holds the largest portion of water, which is salty and unfit for human consumption. Only a small portion of fresh water is usable by humans because around two-thirds of it is frozen in ice caps and glaciers (Bibi *et al.* 2014). Pakistan, currently ranked as the fifth most populous country in the world, used to have a water surplus, but today it has a water deficit. The amount of water per person has declined from 1,299 m^3 in 1996–1997 to 1,100 m^3 in 2006, and by 2025, it is expected to be less than 700 m^3 per person (Murtaza & Zia 2012). According to the World Health Organization reports, the water quality in Pakistan is likewise inadequate. Just 65% of the population has reliable access to sources of improved water. People's perspective of bottled water use is significantly influenced by their education and knowledge about the use of high-quality water. Water quality in Khyber Pakhtunkhwa, a province of Pakistan, is getting increasingly worse daily as a result of poor management and overpopulation (Bibi *et al.* 2014). Many environmental sources are becoming a major problem leading to the pollution of drinking water (Ahmed & Shafique 2019). In underdeveloped nations, contaminated drinking water is frequently seen as a serious health risk with the majority of fatal diseases linked to it. After the Millennium Development Goals era, which came to an end in 2015, the Sustainable Development Goals (SDGs) were endorsed by the United Nations to 'Ensure access to water and sanitation for everyone and ensure sustainable management of water resources' (Zoungrana 2021). According to estimates from the World Health Organization, unsafe water, poor hygiene, and inadequate sanitation cause 1.7 million deaths and 54.2 million disabilities globally each year. Nine out of 10 of these fatalities are child deaths, and almost all of them occur in underdeveloped nations (Jalan *et al.* 2009). An estimated 5 million children die each year in developing nations because of poor water quality (Haydar *et al.* 2009). According to the Government of Pakistan's economic report (2008), there are around 50 million people in Pakistan who lack access to safe drinking water (Zahid 2018). Consuming polluted water is associated, among other things, with microorganisms that cause diarrheal diseases. These pathogens are the primary cause of gastrointestinal infections (Akbar *et al.* 2013). Further, according to community health research, roughly 50% of diseases and 40% of fatalities in Pakistan are caused by poor drinking water quality (Daud *et al.* 2017). A quarter of the population is sickened by contaminated water. According to estimates, three million Pakistanis are affected by waterborne diseases each year, yet 0.1 million of these result in death (Akbar *et al.* 2013). The abundance of surface and groundwater resources in Pakistan is a gift from nature. Water supplies have been severely strained by industrialization, urbanization, and rapid population increase (Daud *et al.* 2017). As many water sources have been contaminated and drained owing to the rapid economic and demographic expansion, now water has become a scarce resource (Ahmad & Sattar 2010). Many trace elements found in drinking water are said to be crucial for human health. Nevertheless, high quantities of these metals, such as zinc (Zn), copper (Cu), iron (Fe), manganese (Mn), cadmium (Cd), nickel (Ni), and lead (Pb), and their excessive use may result in serious health issues (Memon *et al.* 2011). There are two possible ways to ensure that households have ready access to clean water: (1) by supplying treated water at the point of collection and (2) by enabling household water treatment (HWT). In the first approach, the study found that even when water is safe at the source, it is frequently polluted by the time it reaches the point of collection, placing water in frequently supplied systems at significant risk of contamination throughout the distribution process (Pickering *et al.* 2019). Compared with providing treated water at the collection site, HWT is the more efficient approach for providing clean drinking water (Lantagne *et al.* 2006). In the present study, HWT includes boiling water, the addition of chlorine/bleach, straining through cloth, filtration, and water usually treated by solar disinfection used to make water clean and safe. Unfortunately, according to several types of research, despite its beneficial effects the adaptability of HWT is quite low (Luby *et al.* 2000). The adoption of HWT by consumers is influenced by several variables. Past studies found that income (Wedgworth *et al.* 2014), education (Janmaat 2007), education of female household members (Wood *et al.* 2012), age of the household head (du Preez *et al.* 2011), household size (Zabo *et al.* 2016), the income of the household (Fadhullah *et al.* 2022), the locality of residence (Ahmad *et al.* 2019), type of water source (Morakinyo *et al.* 2015), and perception about the quality and usefulness of HWT (Kausar *et al.* 2011) are the primary determinants of HWT adoption. The desire for better water sources in Pakistan is significantly influenced by the household head's education level and the quality of the existing water. Compared with male-headed households, female-headed households had better access to drinking water sources. The degree to which the gender of the family head and better drinking water sources are related, however, is strongly influenced by the income index (Morakinyo *et al.* 2015). Family size and distance of the residence from the water source have a negative influence on the consumption of safe drinking water (Wilson & Boehland 2005). Alvisi *et al.* developed three water level forecasting models: artificial neural networks, the Mamdani fuzzy logic, and the Takagi–Sugeno fuzzy logic. They assessed their performance using different data inputs and found that fuzzy logic models worked well with fewer variables and logic statements, while the artificial neural network model excelled with more detailed information. However, the fuzzy logic models occasionally failed to recognize input combinations during testing, unlike the artificial neural network approach (Alvisi *et al.* 2006). Zhu *et al.* investigated machine learning models, including the feed-forward neural network (FFNN) and deep learning (DL), to predict monthly water levels in 69 temperate lakes in Poland. They found that both models performed well, with FFNN slightly outperforming DL for a considerable number of lakes. These results emphasize the effectiveness of traditional machine learning in lake water level forecasting, with implications for water resources management (Zhu *et al.* 2020). In their study, Abdalrahman *et al.* harnessed the power of artificial neural network techniques, specifically multilayer perceptron (MLP) and Elman neural network (ENN), to develop an optimal model for predicting the infiltration rate (IR) of treated wastewater infiltration basins. By incorporating key water quality parameters, such as total suspended solids (TSS), biological oxygen demand (BOD), electric conductivity (EC), pH, total nitrogen (TN), total phosphorous (TP), and hydraulic loading rate (HLR), their approach represents a significant advancement in the accurate prediction of IR in wastewater treatment processes (Abdalrahman *et al.* 2022). Similarly, another recent study by Ebtehaj *et al.* introduced hybrid models, the generalized structure group method of data handling and the adaptive neuro-fuzzy inference system with fuzzy C-means, for predicting daily water levels at Telom and Bertam stations in Malaysia's Cameron Highlands. These models outperformed traditional approaches, showcasing their effectiveness in daily water level prediction (Ebtehaj *et al.* 2021). Despite several studies addressing water-related challenges and predictions using various techniques, there appears to be a gap in the existing literature. None of the previous research has explored the complex survey data by employing weighted analysis. Given the complex nature of the survey and its potential impact on the study's findings, the use of a weighted descriptive statistics and other statistical models using weights remains unexplored. Therefore, this study aims to fill this gap by applying weighted analysis to examine the consumption of safe drinking water and its determinants, shedding light on previously unaddressed aspects of this critical issue.

The primary objective of our study is to investigate household weighted preferences and determinants of safe drinking water, with a focus on Pakistan. While the study is region-specific, it addresses critical local challenges in ensuring access to safe drinking water. Moreover, Pakistan's significance on the global stage cannot be understated. Pakistan, as the fifth most populous country in the world and one that is also at risk of climate change, faces inevitable water scarcity issues. This makes Pakistan a crucial candidate for water resource research. However, we acknowledge that despite Pakistan's significance, the global impact of our research is limited for several reasons: First, Pakistan's unique cultural, environmental, and socioeconomic aspects may restrict the generalizability of our findings to regions with different conditions and water-related challenges. Global water resource management requires tailored approaches for diverse geographic and environmental conditions. Second, the determinants of safe drinking water consumption identified in Pakistan may not apply universally because of the significant regional variations in climate, water sources, infrastructure, and cultural practices. Third, global water resource management is influenced by numerous complex factors, including climate change, transboundary water agreements, and geopolitics, which our study does not directly address. To significantly impact global water resources, research on a larger scale, encompassing diverse regions and considering the intricate web of global water resource factors, would be necessary. Our study, while valuable for addressing Pakistan's local and national water challenges, has inherent limitations that prevent it from directly influencing global-level decisions. To conclude, we grappled with several notable challenges to comprehensively address the complexities of drinking water preferences and HWT adoption in Pakistan. These included navigating the intricate interplay of urban and rural disparities, diverse socioeconomic factors, and the necessity to incorporate weighted and unweighted analyses, all while striving to provide valuable insights for policymakers and water resource management.

2. MATERIAL AND METHODOLOGY

The Pakistan Demographic and Health Survey (PDHS) 2017–2018 data set has been used in the study. Over the years, DHS has been conducted with funding and assistance from the United States Agency for International Development (USAID). A total of 16,240 households were interviewed for PDHS 2017–2018. A brief description of the data that are used in the analysis is summarized in the following sections.

2.1. Data and sampling scheme

The data used in this study are derived from the Pakistan Demographic and Health Survey (PDHS) conducted in 2017–18. The PDHS provides a representative sample of Pakistan's socioeconomic, demographic, and health characteristics. The results of the 2017–2018 PDHS provide comprehensive and representative data at the national and regional levels, including urban and rural splits. The survey findings accurately represent the four provinces of Punjab, Sindh, Khyber Pakhtunkhwa, and Balochistan, as well as the regions of Azad Jammu and Kashmir (AJK) and Gilgit-Baltistan (GB). Furthermore, the estimates obtained from the 2017–2018 PDHS are also representative of the Islamabad Capital Territory (ICT) and the Federally Administered Tribal Areas (FATA). The survey utilized a two-stage stratified sampling approach to ensure representative coverage. The eight regions were divided into urban and rural areas, resulting in 16 sampling strata. Within each stratum, a two-stage selection process was employed to independently select the samples. In the first stage, 580 clusters were selected using a comprehensive selection process, followed by systematic sampling of households in the subsequent stage. To ensure comprehensive coverage, a household listing operation was conducted across all selected clusters. The households were then carefully chosen using an equal probability systematic selection method, with a fixed number of 28 households per cluster. This resulted in a total sample size of approximately 16,240 households. The sampling frame used in the PDHS 2017–2018 was compiled from the Pakistan General Population and Housing Census conducted in May 2017. The frame included a total of 168,943 enumeration blocks (EBs), of which 55,365 were located in urban areas and 113,578, in rural areas.

2.2. Sampling weights

The majority of large-scale population surveys deviate from simple random sampling methods commonly found in introductory statistics textbooks. Instead, they employ sophisticated sample designs that assign different probabilities of inclusion to the sampling units. When utilizing data obtained from surveys employing such complex sample designs, it is crucial to incorporate the sampling weights into the statistical analyses, including both descriptive and inferential statistics (Winship & Radbill 1994). To achieve an accurate representation of the 2017–2018 PDHS data at both the national and domain levels, it is imperative to employ sampling weights. The reason for this is the non-proportional distribution of the sample across different regions and urban and rural areas. These weights are derived from the selection and inclusion probabilities. Hence, it is important to first elaborate on these probabilities, so let us define that this is the selection probability of the ith cluster at the first stage in stratum h, and P_{2hi} is the selection probability of a household at the second stage within the ith selected cluster. Let b be the number of EBs chosen at stratum h. And let us suppose N_{hi} is the number of households in the ith cluster determined by the sampling frame, and $\sum_{i=1}^{h} N_{hi}$ denotes the total number of households in the stratum. Then the probability of selecting the ith EB in stratum h from the 2017–2018 PDHS sample is estimated as follows (1):

$$P_{(EBhi)} = \frac{b_h N_{hi}}{\sum_{i=1}^{h} N_{hi}} \tag{1}$$

Let us represent the proportion of households in the region to all households in EB i in the stratum h. The sample probability of picking a cluster i in stratum h is defined as follows (2):

$$P_{1hi} = \frac{b_h N_{hi}}{\sum N_{hi}} \times s_{hi} \tag{2}$$

Furthermore, let M_{hi} and express the count of households listed within the cluster and the count of households selected from the cluster, respectively, then the chance of choosing at the second stage is determined as follows (3):

$$P_{2hi} = \frac{d_{hi}}{M_{hi}} \tag{3}$$

As the selection at both stages is independent, the final probability of selection of every household in a cluster i of the stratum h in the PDHS 2017–2018 is thus the product of the selection probabilities defined in Equations (2) and (3) and is computed through the following equation.

$$P_{hi} = P_{1hi} \times P_{2hi} \tag{4}$$

In the stratum h, the sampling weight assigned to each household is calculated as the reciprocal of its overall selection probability, as illustrated in the following equation.

$$W_{hi} = \frac{1}{P_{hi}} \tag{5}$$

A spreadsheet was created to calculate design weights, which took into account various sampling parameters and selection probabilities. These design weights were subsequently adjusted to accommodate non-response at the cluster, household, and individual levels in both the women's and men's surveys. Non-response at the individual level can result in discrepancies between the sampling weights assigned to households and individuals. To ensure consistency, measures were taken to maintain consistency throughout the analyses, and the final sampling weights underwent a normalization process to ensure that the total number of unweighted cases equaled the total number of weighted cases at the national level, both for household weights and individual weights. It is worth noting that the normalized weights used in the analysis are relative, meaning that they are suitable for estimating means, proportions, and ratios. However, they are not appropriate for estimating population totals or combining data. Additionally, the number of weighted cases obtained using normalized weights does not directly indicate the survey's precision, as these weights are relative. This is particularly important to consider for areas that were oversampled.

2.3. Multinomial logistic regression model

Logistic regression is a commonly used statistical method to analyze the relationships between one or more predictor variables (X) and a categorical dependent variable (Y). The dependent variable (Y) can consist of two or more categories, including data with nominal or ordinal measurements. When there are multiple possible values for the dependent variable (Y), logistic regression can be adopted to handle this situation. This modified model is referred to as the multinomial logistic (MNL) regression. Let us have a variable Y containing possible values c, represented as $(0,1,\ldots, c-1)$, and then let us suppose that we have p independent variables, which are also denoted as covariates $\mathbf{x} = (x_1, x_2, \ldots, x_p)$. According to a certain value of x, the multinomial probability of each result is

$$P(Y = 0/x) = \frac{1}{1 + e^{g_1(x)} + \ldots + e^{gc-1(x)}} \tag{6}$$

$$P(Y = 1/x) = \frac{e^{g_1(x)}}{1 + e^{g_1(x)} + \ldots + e^{gc-1(x)}} \tag{7}$$

$$P(Y = c - 1/x) = \frac{e^{g_{c-1}(x)}}{1 + e^{g_1(x)} + \ldots + e^{gc-1(x)}} \tag{8}$$

As a result, the logit of group j relative to the baseline group is as follows:

$$g_j(x) = \ln\left[\frac{P(Y = j/x)}{P(Y = 0/x)}\right] = \beta_{j0} + \beta_{j1}x_1 + \ldots + \beta_{jp}x_p \tag{9}$$

where $j = 1, \ldots, c - 1$.

2.4. Dependent variables

This study involves two outcome variables, the source of drinking water and the methods of treating water in households. Both of these variables are defined in the following subsections.

2.4.1. Source of drinking water

In the survey, 17 sources of drinking water were considered. We grouped these 17 sources into 6 different water sources, as follows: (1) bottled/filtered water, (2) piped water, (3) protected well, (4) unprotected well, (5) surface water, and (6) bought water from commercial entities.

2.4.2. Household water treatment

We established a binary variable to indicate the adoption of purification methods within households. Three sources are used in the PDHS: (1) boiling, (2) chlorine/bleach treatment, and (3) water filtration. Therefore, we created a binary variable from these sources. It takes a value of 1 when any form of purification method is implemented at home, and 0 when no purification method is employed by the household.

2.5. Independent variables

This study involves eight independent variables, which are described as follows.

2.5.1. Age of the household head

Over 1.8 million children under the age of five die from diarrhea every year, a condition that is both preventable and curable (du Preez *et al.* 2011). It is hypothesized that households with older heads are less likely to use safe drinking water or adopt modern purifying methods. Different age categories are used in the analysis of safe drinking water in Pakistan. The four age categories are 15–25, 26–39, 40–59, and 60 + .

2.5.2. Level of education of household head

Education has a detrimental impact on the choice to treat drinking water, but it may have a positive impact on the degree of treatment for those who do (Janmaat 2007). In the data set, the variable 'Education' is divided into four categories: no education, primary, secondary, and higher education. We hypothesize that education will positively affect the role of safe drinking water sources and the use of purifying methods.

2.5.3. Family size of the household

Family size became an important consideration when choosing the drinking water source for the household due to the significant results for all the alternatives. It is commonly known that household size has a favorable impact on water usage (Zabo *et al.* 2016). We hypothesize that, as the size of the household increases, it will reduce the chances of using filtered water as well as adopting HWT methods. In the data set, the minimum household size is 1 and the maximum household size is 44 members. The variable used in the data set is categorized as 1–5, 6–10, 11–15, and 16 or more members.

2.5.4. Wealth of household

In Pakistan, there is a belief that individuals in lower socioeconomic levels are only marginally ready to pay because they feel that water should be supplied for free since it is a natural resource. Residents eventually do pay a price for availing water provided through the piped infrastructure, either by extracting water manually, traveling great distances to collect it, or buying water from nearby vendors or in bottled form (Wedgworth *et al.* 2014). Household wealth affluence can simultaneously have a direct and indirect impact on the drinking water quality in a household (Fadhullah *et al.* 2022). The wealth index was used in the PDHS 2017–2018 to describe a household's wealth. There are around 40 different asset variables used to compute the wealth index of a household. In the computation process, principal component analysis (PCA) was used. With the help of PCA, we categorized the 40 different asset variables, including the housing facilities and other materials, into five different categories. These five wealth index categories may each have a score from 1 to 5, where 1 indicates the poorest, 2 indicates the poor, 3 shows the middle class, 4 indicates the richer, and 5 indicates the richest. Therefore, we hypothesize that the increase in wealth of a household could increase the chance of bottled/filtered water use and the adoption of HWT methods.

2.5.5. Exposure to media

Exposure to media contains three variables: (1) reading the newspaper, (2) watching TV, and (3) listening to the radio. Therefore, we construct, with the help of STATA software, a binary variable that is 'exposure to media' from these three variables, which take the value 1 if a household either reads the newspaper, watches TV, or listens to the radio and take the value 0 otherwise. It is hypothesized that a household that has media exposure will likely increase the probability of using bottled/filtered water and could adopt better household water treatments.

2.5.6. Women's empowerment

Women's empowerment contains several aspects. A few of them are household decision-making and economic contribution toward the household, control over resources, and freedom of moment. In the data set, the variable women's empowerment contains five responses, as follows: (1) respondent alone, (2) respondent and husband/partner, (3) husband/partner alone, (4) family elders, and (5) others. Therefore, we make a binary variable, with the first two responses assigned the value 1, which means that the women have autonomy, while the last three responses are assigned the value 0, which means that the women have no autonomy in decision-making. The study hypothesizes that women's empowerment will increase the chance of bottled/filtered water use and the adoption of HWT.

2.5.7. Distance to the water source

We utilized the information on walking distance to measure the relative distance to the water source in the data set. In the PDHS data set, the variable has three categories: (1) at home, (2) it takes up to 15 min to reach the water source, and (3) it takes more than 15 min to reach the water source. The study hypothesis is that the greater the distance to the water source, the lesser the chances of using bottled/filtered water or adopting HWT.

2.5.8. Region

Rural and urban areas are the two options available in the data set. It is hypothesized that households in urban areas are more likely to use bottled/filtered water and to adopt HWT. Therefore, we created a binary variable assigning a value of 1 if a household is present in the rural areas, while assigning 0 if a household is present in the urban areas.

3. RESULTS AND DISCUSSION

To assess the impact of weighting on our study's findings, we present a comparison of both weighted and unweighted descriptive statistics for the key variables of interest. Weighting is a statistical technique used to account for the complex sampling design and ensure that the sample accurately represents the population. The weighted estimates are adjusted to give greater influence to underrepresented groups and provide a more accurate reflection of the entire population. We employ a *t*-test to determine whether significant differences exist between the weighted and unweighted estimates. This analysis helps us evaluate the potential bias introduced by the sampling and weighting methods and allows us to make informed conclusions regarding the effects of these adjustments on our study's results. Table 1 provides the weighted and unweighted descriptive statistics, along with the results of the *t*-test, shedding light on the impact of weighting on our findings.

In Table 1, the utilization of survey weights within the context of our research reveals significant variations in key variables, underscoring the critical role of these weights in providing a more accurate depiction of our complex survey data. Notably, the 7.38% difference between urban and rural populations becomes more pronounced when considering survey weights, reflecting the complex nature of our study area. The impact of survey weights is especially evident in the realm of water source preferences, where the differences extend to 3.97% for bottled/filtered water, 9.85% for piped water, and 1.61% for protected wells. These variations are pivotal in understanding the dynamics of water access, where survey weights enhance our comprehension of disparities.

Moreover, the adoption of HWT reveals a 3.49% difference between 'No' and 'Yes' categories, emphasizing the importance of accounting for survey weights in assessing practices that can significantly impact public health. When examining distance to water sources, the application of survey weights illuminates substantial differences: a –16.55% disparity for 'At home,' a −6.35% disparity for 'Up to 15 min,' and a −10.2% disparity for 'Above 15 min,' offering a more nuanced view of accessibility. In the context of age distribution, the 60+ age group displays a 3.35% difference, while the 26–39 age group exhibits a −2.76% difference when survey weights are considered. These distinctions are essential in capturing the diverse demographic landscape accurately. The influence of survey weights also extends to variables, such as household size, education, wealth, media exposure, and women's empowerment, where they highlight variations that carry profound implications for policy and program development. In summary, the table elucidates the crucial role of survey weights in uncovering meaningful disparities and providing a more insightful understanding of our complex survey data.

Table 2 offers a comprehensive analysis of χ^2 tests investigating the water source variable, revealing noteworthy variations in χ^2 values and *p*-values when comparing unweighted and weighted data for each variable. For 'Location,' both unweighted and weighted data exhibit a significant relationship with water source preferences in terms of location, with *p*-values of 0.000. Interestingly, the weighted data significantly increase the χ^2 value, highlighting a stronger association when survey weights are considered. For 'Age of the Household Head,' both unweighted and weighted results reveal significant associations with water source preferences (*p*-value = 0.000). The application of survey weights remarkably amplifies the χ^2 values, emphasizing the heightened significance of these age-related effects. When assessing 'household size,' it is evident that the χ^2 values and *p*-values differ between unweighted and weighted data. While the unweighted data demonstrate a significant relationship with water source choices (*p*-value = 0.000), the weighted data significantly increase the *p*-value to 0.488, indicating a loss of significance. This underscores the critical role of survey weights in accurately capturing variations related to household size and water source preferences and highlights the importance of recognizing cases where significance may diminish in the weighted analysis. χ^2 tests for 'Media Exposure' display significant connections with 'water source' in both unweighted and weighted data (*p*-value = 0.000). Here, the weighted data notably increase the χ^2 values, highlighting the profound impact of media exposure when survey weights are applied. For 'Education,' significant differences in water source choices are evident across education levels in both unweighted and weighted data, with *p*-values of 0.000. However, the weighted data notably augment the χ^2 values, emphasizing the importance of survey weights in accurately capturing these educational disparities. Both unweighted and weighted estimates reveal significant associations between 'women's empowerment' and 'water source' (*p*-value = 0.000). The application of survey weights distinctly increases the χ^2 values, underscoring the substantial influence of women's empowerment on water source choices. In the last variable for 'Wealth of the Household' for 'water source,' the χ^2 tests display significant relationships between both unweighted and weighted data (*p*-value = 0.000). Nevertheless, the weighted data significantly raises the χ^2 values, highlighting the critical role of survey weights in understanding the economic aspects of water source preferences. These nuanced variations emphasize the intricate associations uncovered by survey weights, underscoring their pivotal role in assessing the complex relationships between these variables and water source preferences in the context of a complex survey design.

Table 3 presents the outcomes of χ^2 tests examining the relationships between water source preferences and various variables in both unweighted and weighted data, highlighting essential distinctions in χ^2 values and *p*-values that bear substantial implications for real-world scenarios. In the context of 'Location,' unweighted data initially show no significant relationship with water source choices (*p*-value = 0.685), implying that the choice of water source does not substantially vary between urban and rural areas when survey weights are not considered. However, when survey weights are applied, the relationship becomes highly significant (*p*-value = 0.75), suggesting that survey weights reveal significant disparities in water source preferences between urban and rural areas. This underscores the critical role of survey weights in uncovering location-specific

Table 1 | Descriptive statistics and impact of weights

Variables	Unweighted (%)	Weighted (%)	Differences (%)	*P*-value
Location				
Urban	45.67	53.05	7.38	<0.0001
Rural	54.33	45.95	−7.38	<0.0001
Water source				
Bottled/filtered water	5.56	1.59	−3.97	< 0.0001
Piped water	36.74	46.59	9.85	< 0.0001
Protected well	42.86	44.47	1.61	< 0.0001
Unprotected well	2.14	0.95	−1.19	< 0.0001
Surface water	10.47	6.18	−4.29	< 0.0001
Bought water from commercial entities	2.23	0.22	− 2.11	< 0.0001
Adoption of HWT				
No	91.42	94.31	3.49	< 0.0001
Yes	8.58	5.69	− 2.89	< 0.0001
Distance to water source				
At home	71.02	87.57	16.55	< 0.0001
Up to 15 min	12.38	6.03	− 6.35	< 0.0001
Above 15 min	16.60	6.40	− 10.2	< 0.0001
Age of household head				
15–25	2.30	2.05	− 0.25	< 0.0001
26–39	26.45	23.69	− 2.76	< 0.0001
40–59	54.00	53.66	− 0.34	0.5418
60 +	17.25	20.60	3.35	< 0.0001
Household size				
1–5	18.12	16.04	− 2.08	< 0.0001
6–10	57.99	58.03	0.04	0.9442
11–15	16.25	15.43	− 0.82	0.0433
16 +	7.64	10.50	2.86	< 0.0001
Education				
No education	60.79	65.46	4.67	< 0.0001
Primary education	13.56	11.76	− 1.8	< 0.0001
Secondary education	16.38	14.98	− 1.4	0.0005
Higher education	9.27	7.79	− 1.48	< 0.0001
Wealth				
Poorest	28.92	19.32	− 9.6	< 0.0001
Poorer	22.96	23.21	0.25	0.5961
Middle	18.75	22.17	3.42	< 0.0001
Richer	15.68	21.50	5.82	< 0.0001
Richest	13.68	13.80	0.12	0.7565
Media exposure				
No	34.99	40.52	5.53	< 0.0001
Yes	65.01	59.48	− 5.53	< 0.0001

(*Continued.*)

Table 1 | Continued

Variables	Unweighted (%)	Weighted (%)	Differences (%)	*P*-value
Women's empowerment				
No	55.07	51.58	– 3.49	< 0.0001
Yes	44.93	48.42	3.49	< 0.0001

Table 2 | Summary of unweighted and weighted χ^2 tests for water source

	Unweighted			Weighted		
Variables	χ^2	df	*P*-value	χ^2	df	*P*-value
Location	4,500	5	0.000	1,232.89	5	0.000
Age of the household head	305.11	15	0.000	100.10	15	0.333
Household size	261.28	15	0.000	100	15	0.488
Media exposure	2,700	5	0.000	1,148.27	5	0.000
Education	3,700	15	0.000	1,472.79	15	0.000
Women's empowerment	860.12	5	0.000	349.54	5	0.000
Wealth of the household	5,000	20	0.000	2,034.84	20	0.000

Table 3 | Summary of unweighted and weighted χ^2 tests for household water treatment

	Unweighted			*m*		
	χ^2	Df	*P*-value	χ^2	Df	*P*-value
Location	0.1642	1	0.685	1.51	1	0.75
Age of the household head	18.04	3	0.000	22.80	3	0.19
Household size	35.40	3	0.000	7.97	3	0.77
Media exposure	53.76	1	0.000	18.11	1	0.13
Education	270.82	3	0.000	46.93	3	0.07
Women's empowerment	36.07	1	0.000	0.81	1	0.73
Wealth of the household	35.74	4	0.000	40.47	4	0.19

nuances in real-life water source preferences. For 'Age of the Household Head,' unweighted data indicate a highly significant relationship with water source preferences (p-value = 0.000), implying that the age of the household head significantly impacts water source choices. In contrast, when survey weights are applied, the p-value increases to 0.19, rendering the relationship insignificant. This suggests that age may not be a critical factor in the context of survey weights, highlighting the nuanced influence of age on water source preferences in real-life scenarios. When considering 'Household Size,' unweighted data demonstrate a significant relationship with water source preferences (p-value = 0.000), suggesting that household size significantly influences these choices. However, the application of survey weights reduces the significance, with a p-value of 0.77, implying that the significance diminishes in real-world contexts when survey weights are considered. This emphasizes the importance of recognizing cases where significance may wane in the presence of survey weights. In the analysis of variables such as 'Media Exposure,' 'Education,' 'Women's Empowerment,' and 'Wealth of the Household,' it is evident that while the unweighted data indicate highly significant relationships with water source preferences (p-value = 0.000), the significance decreases when survey weights are applied (e.g., p-value = 0.13 for 'Media Exposure'). This implies that when accounting for survey weights, these variables may not have as pronounced an impact on water source choices as initially observed without the consideration of weights. These findings underline the importance of scrutinizing the influence of these factors in the presence of survey weights in real-life contexts, where their impact may be less significant than initially assumed based on unweighted data. Such nuanced insights are invaluable for making informed decisions and policy recommendations.

Table 4 presents the results of the unweighted MNL model, analyzing the influence of various demographic variables on the choice of drinking water source categories: 'Bottled/filtered Water,' 'Piped Water,' 'Protected Well,' 'Unprotected Well,' 'Surface Water,' and 'Bought Water from Commercial Entities.'

In the unweighted analysis, it is evident that specific age groups exhibit significant associations with particular water source preferences. For instance, individuals aged 26–39, 40–59, and 60+ show heightened likelihood of selecting 'Protected Well' as a source of water while being less inclined toward 'Bottled/filtered Water,' 'Piped Water,' and 'Surface Water.' Moreover,

Table 4 | Summary of unweighted (MNL) for the determinants of drinking water sources

	Water sources					
Variables	**Bottle/filtered water**	**Piped water**	**Protected well**	**Unprotected well**	**Surface water**	**Bought water from commercial**
Region						
Urban	1					
Rural		0.689*	1.453*	2.640*	2.557*	−0.22*
Age of the household head						
15–25	1					
26–39		0.144	0.056	0.982*	−0.085	−0.82*
40–59		0.274	−0.112	0.784*	0.194	−1.01*
60 +		0.231	−0.157	0.828*	0.272	−1.11*
Household size						
1–5	1					
6–10		0.182*	0.170*	0.519*	0.054	0.360*
11–15		−0.052	−0.138	0.379*	−0.328*	0.399*
16 +		−0.033	0.223*	0.942*	−0.074	0.806*
Media exposure						
No	1					
Yes		−1.05*	−1.09*	−1.82*	−1.10*	−1.24*
Education						
No education	1					
Primary		−0.726*	−0.670*	−1.36*	−0.283*	−1.32*
Secondary		−0.655*	−0.764*	−0.66*	0.222*	−1.426*
Higher		−0.882*	−1.33*	−1.60*	−0.106	−1.41*
Women's empowerment						
No	1					
Yes		−0.503*	−0.226*	−0.995*	−0.563*	−0.77*
Wealth of the household						
Poorest	1					
Poorer		0.286*	0.042	−0.416*	−0.509*	−0.27*
Middle		−0.159	−0.278*	−1.96*	−1.40*	−0.86*
Richer		−0.46*	−0.507*	−2.72*	−2.38*	−0.99*
Richest		−1.62*	−1.33*	−3.32*	−4.113*	−1.73*
LR χ^2 (80) = 11,800.78						
Probability > χ^2 = 0.000						
Pseudo-R^2 = 0.0934						

The significance level of * is 0.05.

household size plays a notable role, with larger households being more likely to opt for 'Bottled/filtered Water,' 'Piped Water,' and 'Protected Well.' Education is another significant factor, as those with higher education tend to choose different water sources. Additionally, media exposure is associated with a reduced likelihood of using any of the examined water sources.

Table 5 portrays the outcomes of the weighted MNL model, emphasizing the impact of demographic variables on water source preferences.

In the weighted analysis, the associations between age groups and water source preferences are consistent with that of the unweighted analysis, with certain age groups favoring 'Protected Well' and avoiding other sources. The influence of

Table 5 | Summary of weighted (MNL) for the determinant of drinking water source

	Water sources					
Variables	**Bottle/filtered water**	**Piped water**	**Protected well**	**Unprotected well**	**Surface water**	**Bought water from commercial**
Region						
Urban	1					
Rural		0.419*	1.57*	3.68*	3.62*	−1.08*
Age of the household head						
15–25	1					
26–39		0.440	0.077	1.37*	−0.103	−0.508
40–59		0.555*	−0.162	0.74	−0.23	−0.82*
60 +		0.47*	−0.126	0.719	0.174	−1.16*
Household size						
1–5	1					
6–10		0.036	0.171*	0.82*	0.197*	0.306*
11–15		−0.164*	−0.109	0.944*	0.120	0.40*
16 +		−0.78*	−0.36*	2.00*	−0.53*	0.293
Media exposure						
No	1					
Yes		−0.91*	−0.90*	−1.51*	−1.16*	−1.25*
Education						
No education	1					
Primary		−0.682*	−0.649*	−2.05*	−0.72*	−1.53*
Secondary		−0.693*	−0.667*	−0.633*	−0.741*	−2.22*
Higher		−0.870*	−1.13*	−2.80*	−0.951*	−1.61*
Women's empowerment						
No	1					
Yes		−0.52*	−0.28*	−1.51*	−1.31*	−1.20*
Wealth of the household						
Poorest	1					
Poorer		0.885*	0.538*	−0.094	0.535*	−0.025
Middle		0.003	−0.386*	−2.22*	−1.59*	−0.83*
Richer		0.002	−0.378*	−2.53*	−1.71*	−0.647*
Richest		−1.26*	−1.47*	−5.02*	−3.30*	−1.79*
LR χ^2 (80) = 10,884.95						
Probability > χ^2 = 0.0000						
Pseudo-R^2 = 0.1241						

The significance level of * is 0.05.

household size is maintained, and education remains a determinant of water source choice. Media exposure continues to reduce the likelihood of using these water sources.

The discrepancies between the unweighted and weighted analyses can be attributed to the complex survey design, as the weighted analysis takes into account the different sizes of households in the study, providing a more precise estimation of population parameters. However, it is essential to note that the practical significance of these associations should be carefully considered.

While both unweighted and weighted analyses reveal statistical significance, the practical importance may vary. The weighted analysis considers the complexity of the survey data, highlighting the real-world implications of these demographic factors on water source choices more accurately. In practical terms, these findings suggest that specific age groups, household sizes, educational levels, media exposure, and demographic disparities indeed play crucial roles in shaping water source preferences. Policymakers and water resource planners can use these insights to tailor interventions and resources more effectively, promoting safe and accessible drinking water sources.

Tables 4 and 5 provide the outcomes of the unweighted and weighted analyses examining the connection between distinct age groups and different water sources. The findings highlight noteworthy correlations between age categories and water source preferences. For individuals aged 26–39, 40–59, and 60 and above, significant positive correlations with protected wells and notable negative correlations with bottled/filtered water, piped water, and surface water emerge in both the unweighted and weighted analyses. This implies a heightened likelihood of the opting of protected wells, and a diminished likelihood of the selecting of bottled/filtered water, piped water, and surface water, by individuals within these age ranges. People of all ages are more likely to use protected wells than other water sources. This trend is especially pronounced among people aged 26–39, 40–59, and 60 and above.

Education is categorized into four groups: no education, primary education, secondary education, and higher education. People with higher education are less likely to use bottled/filtered water, piped water, protected wells, surface water, and commercial water as their sources of drinking water. This is also true for people with primary and secondary education.

The variable 'Household size' used in the data set is categorized as 1–5, 6–10, 11–15, and 16 or more members. Household size is a significant factor in the selection of drinking water sources. Larger households are more likely to use bottled/filtered water, piped water, and protected wells. This is because larger households have a greater water demand and are more likely to be concerned about the quality of their water.

As mentioned previously, the study used a wealth index to measure household economic status. The wealth index was calculated using PCA, which is a statistical technique that combines multiple variables into a single index. The wealth index was divided into five categories, from poorest to richest. The study found that household wealth is associated with the selection of water sources. The poorer households are more likely to use bottled/filtered water and piped water, while middle-income and richer households are more likely to use protected wells. The richest households tend to have the lowest likelihood of using any of the water sources.

The study found that people who were exposed to media were less likely to use any of the water sources examined. This is likely because media exposure can increase awareness of the risks of waterborne diseases, such as cholera, dysentery, and typhoid fever. People who are aware of these risks may be more likely to choose safer water sources, such as bottled/filtered water or water that has been boiled. The study also found that the weighted analysis provided a more accurate understanding of the relationship between media exposure and water source selection. This is because the weighted analysis took into account the different sizes of the households in the study.

Within the data set the variable 'women's empowerment' comprises five response options: (1) respondent alone, (2) respondent and husband/partner, (3) husband/partner alone, (4) family elders, and (5) others. To analyze this variable, we constructed a binary variable that assigns a value of 1 to the first two responses, indicating women's autonomy, and a value of 0 to the last three responses, signifying a lack of autonomy. The study found that women's empowerment is associated with the use of specific water sources. Empowered women are more likely to use protected wells and surface water, and less likely to use piped water. There is no significant association between women's empowerment and bottled/filtered water or unprotected wells.

Furthermore, as mentioned previously, the study used a binary variable to measure the region. The variable was coded as 1 if the household was located in a rural area and 0 if it was located in an urban area. The study also used unweighted and weighted data analyses to account for the different sizes of households in the study. The study found that the selection of water sources is influenced by region. Rural areas are more likely to use bottled/filtered water, piped water, protected wells, and unprotected wells than urban areas.

3.1. Determinants of HWT

In the unweighted model as shown in Table 6, the LR χ^2 value is 643.36 with 23 degrees of freedom, showing a statistically significant association between the independent variables and HWT. The p-value of 0.000 confirms the model's significance. The pseudo-R^2 value of 0.0100 indicates that the independent variables account for about 1% of the variation in HWT.

The weighted model as shown in Table 6 also reveals a significant association between the independent variables and HWT, with an LR χ^2 value of 192.19 and 23 degrees of freedom. The p-value of 0.000 further supports the model's significance. The pseudo-R^2 value of 0.0100 indicates that the independent variables account for about 1% of the variation in HWT, consistent with the unweighted model. The study found that there is a significant association between the independent variables and HWT, as evidenced by the significant LR χ^2 values and low p-values. However, the pseudo-R^2 values suggest that the independent variables only explain a small portion of the variation in HWT.

Table 6 delves into the relationship between various factors and HWT practices, considering both unweighted and weighted logistic regression models to account for the sample design. The findings indicate that older household heads (age 40–59 and 60+) are more inclined to engage in water treatment practices compared with their younger counterparts (age 15–25). This age-related association remains consistent in both unweighted and weighted analyses. Education level plays a pivotal role in water treatment practices, with households with members having primary, secondary, and higher education displaying a greater likelihood of treating water compared with those without education. This education-related correlation persists in both unweighted and weighted models. In terms of household size, a significant association is observed with water treatment in the unweighted model. However, this association vanishes when considering the sample design in the weighted model, highlighting the importance of accounting for survey weights. Household wealth exhibits a negative relationship with water treatment, indicating that as household wealth increases, water treatment practices decrease. This inverse association holds true for both unweighted and weighted models. Media exposure reveals a positive correlation with water treatment practices in the unweighted model. Still, this association loses significance when considering the survey weights in the weighted model. The study also explores the link between women's empowerment in household purchases and water treatment. In the unweighted model, there is a positive connection, but this association becomes insignificant when survey weights are applied in the weighted model. Additionally, the study analyzes the impact of the distance to the water source on water treatment practices. Results indicate that households located within 15 min of a water source are more likely to treat their water than those with a water source at home. This association remains statistically significant in both unweighted and weighted models. Finally, the study addresses the urban–rural divide, revealing that households in rural regions are more likely to treat their water than those in urban areas. This association is statistically significant in both the unweighted and weighted analyses. The incorporation of survey weights helps provide a more accurate representation of these associations within the complex survey design, highlighting the need to consider the sampling strategy when drawing meaningful conclusions.

Figure 1 illustrates the unweighted percentages of drinking water sources, categorized by urban and rural areas. It provides valuable insights into the distribution of these water sources within each distinct region. In urban areas, 'Piped Water' emerges as the predominant source of drinking water, constituting 45.79% of the total supply. 'Protected Wells' also hold a substantial place, contributing to 35.17% of the urban population's access to clean water. 'Bottled/Filtered Water' sources make up a notable 9.33%, while 'Surface Water' and 'Commercial Entities' account for smaller yet significant portions at 5.37 and 3.64%, respectively. 'Unprotected Wells' represent a minor fraction at 0.70%. Conversely, rural areas exhibit a distinct pattern. 'Protected Wells' take center stage as the primary sources of drinking water, serving 49.32% of the rural population. 'Piped Water' sources, on the other hand, play a notably diminished role in rural regions, providing access to 29.14%. 'Surface Water' assumes a relatively more significant position at 14.75%, followed by 'Commercial Entities' at 1.04%. 'Bottled/Filtered Water' and 'Unprotected Wells' sources are less common in rural settings, comprising 2.39 and 3.36%, respectively.

Figure 2 presents weighted percentages of drinking water sources, categorized by area and differentiated between rural and urban regions. These percentages offer a valuable overview of how these water sources are distributed in each specific region.

In rural areas, the predominant source of drinking water is 'Protected Wells', accounting for a substantial majority at 65.01%. 'Piped Water' sources, while present, play a somewhat smaller role, providing access to 22.57% of the rural

Table 6 | Estimation result of unweighted and weighted logit model of the in-house water treatment to treat water

Variables	Unweighted		Weighted	
Region	Odd ratio	*P*-value	Odd ratio	*P*-value
Urban	1		1	
Rural	1.248*	0.000	1.303*	0.000
Age of household head				
15–25	1		1	
26–39	1.166	0.278	1.190	0.341
40–59	1.357*	0.030	1.465*	0.035
60 +	1.396*	0.022	1.447*	0.048
Household size				
1–5	1		1	
6–10	0.892*	0.009	1.012	0.817
11–15	0.923	0.170	1.173*	0.022
16 +	0.779*	0.002	0.955	0.656
Education				
No education	1		1	
Primary education	1.455*	0.000	1.204*	0.002
Secondary education	1.660*	0.000	1.510*	0.000
Higher education	1.977*	0.000	2.013*	0.000
Wealth				
Poorest	1		1	
Poorer	0.823*	0.000	0.862*	0.023
Middle	0.883*	0.019	0.905	0.136
Richer	0.883*	0.004	0.689*	0.000
Richest	0.810*	0.001	0.686*	0.000
Media exposure				
No	1		1	
Yes	1.225*	0.000	1.033	0.520
Distance to water source				
At home	1		1	
Up to 15 min	1.633*	0.000	1.251*	0.000
Above 15 min	1.191*	0.001	1.179*	0.008
Women's empowerment in household purchases				
NO	1		1	
Yes	1.125*	0.001	0.942	0.163
Water source				
Bottled water	1		1	
Piped water	1.116	0.137	0.921	0.340
Protected well	0.941	0.411	0.961	0.631
Unprotected well	0.966	0.812	1.260	0.195
Surface water	1.615*	0.000	0.391*	0.000
Bought water from commercial entities	0.836	0.191	0.958	0.828

(*Continued.*)

Table 6 | Continued

Variables Region	Unweighted Odd ratio	P-value	Weighted Odd ratio	P-value
LR χ^2 (23) = 643.36		LR χ^2 (23) = 192.19		
Probability > χ^2 = 0.000		Probability > χ^2 = 0.0000		
Pseudo-R^2 = 0.0225		Pseudo-R^2 = 0.0100		

*$p < 0.05$.

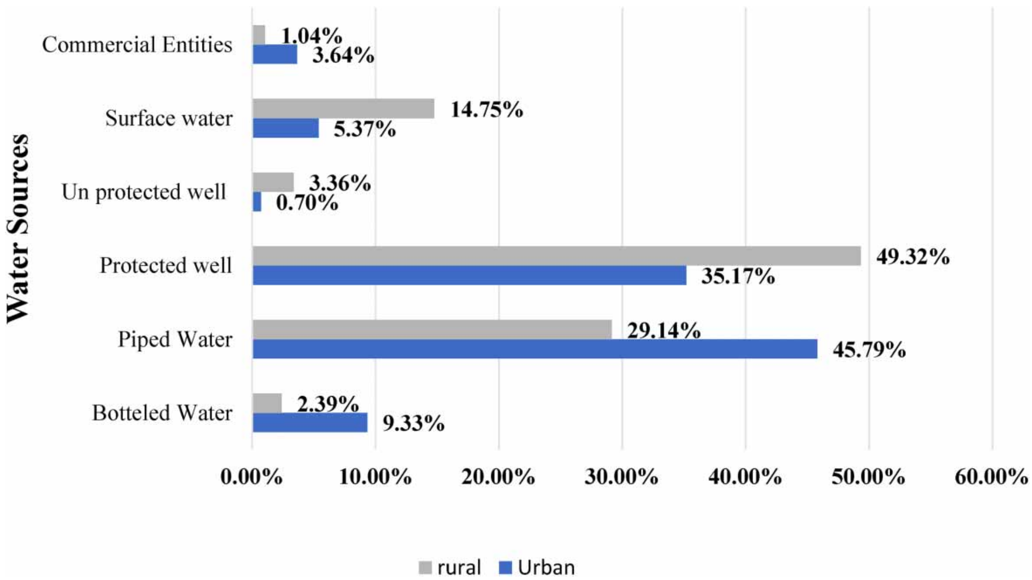


Figure 1 | Unweighted percentages of drinking water sources with respect to area.

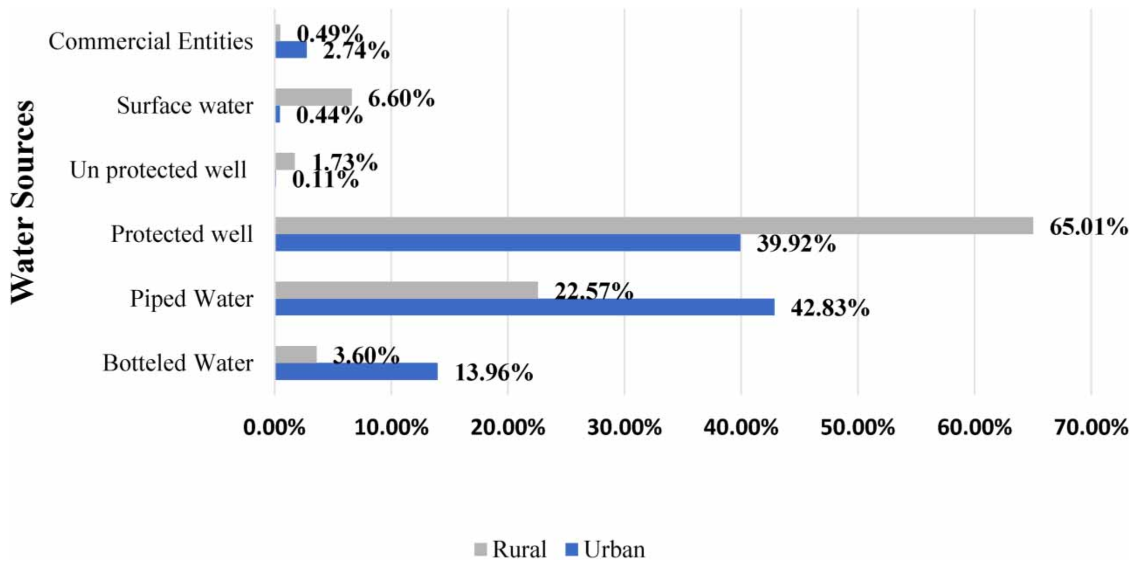


Figure 2 | Weighted percentages of drinking water sources with respect to area.

population. 'Surface Water' represents a relatively significant source at 6.60%, whereas 'Bottled/Filtered Water' sources and 'Unprotected Wells' are less frequently utilized, making up 3.60 and 1.73%, respectively. 'Commercial Entities' contribute to a minor extent, comprising only 0.49% of the total.

In contrast, urban areas exhibit a different pattern. 'Piped Water' sources dominate, providing drinking water to 42.83% of the urban population. 'Protected Wells' remain sources, albeit with a reduced percentage at 39.92%. 'Bottled/Filtered Water' sources are more widely adopted in urban regions, accounting for 13.96%. 'Surface Water' and 'Commercial Entities' represent smaller portions, at 0.44 and 2.74%, respectively. 'Unprotected Wells' are the least utilized sources in urban areas, comprising just 0.11% of the total.

Figure 3 offers unweighted percentages illustrating the distribution of various drinking water sources across different regions. It provides valuable insights into the regional disparities in use of water sources.

In the Punjab region, 'Protected Wells' emerge as the primary source of drinking water, constituting a significant 65.80% of the total. 'Piped Water' sources are also substantial, providing access to 22.33% of the population. 'Bottled/filtered water'

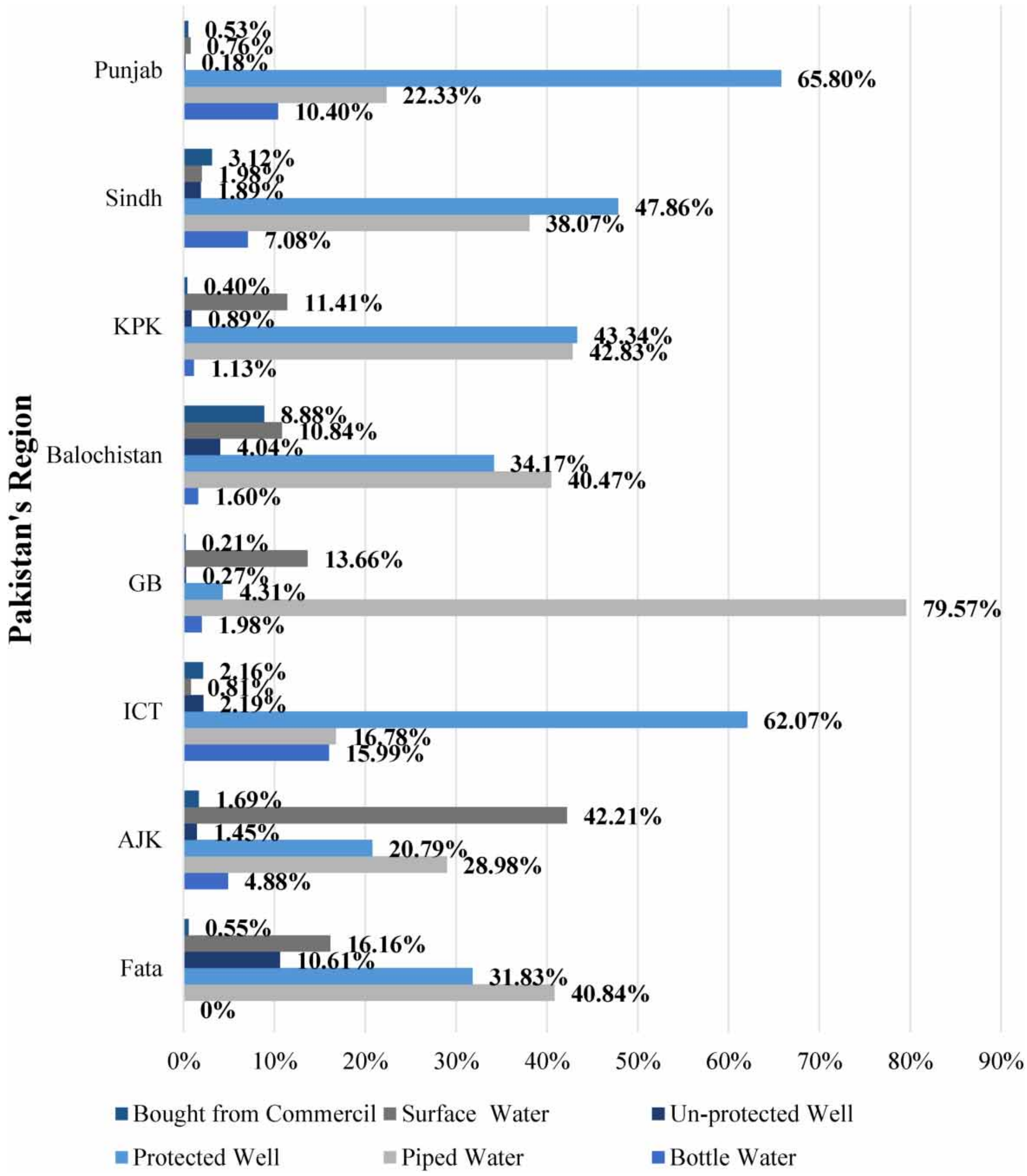


Figure 3 | Unweighted percentages of drinking water sources.

sources account for 10.40%, while 'Surface Water', 'Commercial Entities', and 'Unprotected Wells' represent smaller proportions at 0.76, 0.53, and 0.18%, respectively.

Turning our attention to Sindh, 'Protected Wells' continue to play pivotal roles, with a share of 47.86%. 'Piped Water' sources are also noteworthy, contributing to 38.07% of the population's drinking water. 'Bottled/filtered water' sources make a 7.08% contribution, while 'Unprotected Wells', 'Surface Water', and 'Commercial' sources constitute smaller segments at 1.89, 1.98, and 3.12%, respectively.

In KP (Khyber Pakhtunkhwa), 'Piped Water' emerges as the dominant source, providing drinking water to 42.83% of the region's population. 'Protected Wells' closely follow at 43.34%. 'Surface Water' plays a notable role with an 11.41% contribution, while 'Bottled/filtered water', 'Unprotected Wells', and 'Commercial' sources are relatively less significant at 1.13, 0.89, and 0.40%, respectively.

Balochistan's water source distribution highlights the significance of 'Piped Water', accounting for 40.47% of the total drinking water. 'Protected Wells' contribute 34.17%, while 'Surface Water' makes up 10.84%. 'Bottled/filtered Water', 'Unprotected Wells', and 'Commercial' sources hold smaller portions at 1.60, 4.04, and 8.88%, respectively.

In Gilgit-Baltistan (GB), 'Piped Water' takes center stage as the primary source, with a remarkable 79.57% share. 'Protected Wells' play a minor role at 4.31%, while 'Surface Water' is relatively significant at 13.66%. 'Un-protected Wells', 'Bottled/filtered Water', and 'Commercial Entities' sources make smaller contributions at 0.27, 1.98, and 0.21%, respectively.

The ICT showcases 'Protected Wells' as the primary source, contributing 62.07% of the drinking water. 'Bottled/filtered Water' sources are substantial, accounting for 15.99%. 'Piped Water' contributes 16.78%, while 'Surface Water', 'Unprotected Wells', and 'Commercial' sources have smaller shares at 0.81, 2.19, and 2.16%, respectively.

In Azad Jammu and Kashmir (AJK), 'Surface Water' takes the lead as the dominant source, providing 42.21% of the drinking water. 'Piped Water' is also substantial at 28.98%. 'Protected Wells' contribute 20.79%, while 'Unprotected Wells', 'Bottled/filtered Water', and 'Commercial Entities' sources hold smaller percentages at 1.45, 4.88, and 1.69%, respectively.

Lastly, FATA (Federally Administered Tribal Areas) primarily relies on 'Piped Water', constituting 40.84% of the drinking water. 'Protected Wells' make a 31.83% contribution, while 'Surface Water' is notable at 16.16%. 'Un-protected Wells', 'Bottled/filtered Water', and 'Commercial Entities' sources have smaller roles, with 'Commercial' being the least utilized at 0.55%.

Figure 4 provides weighted percentages, offering insight into the distribution of various drinking water sources across diverse regions. It highlights the prevalence of these sources in each specific area and underscores regional disparities.

In the Punjab region, 'Protected Wells' emerge as the primary source of drinking water, accounting for a substantial majority at 66.03%. 'Piped Water' sources are also significant, providing access to 22.45% of the population. 'Bottled/filtered Water' sources constitute 10.55%, while 'Surface Water', 'Commercial Entities', and 'Unprotected Wells' represent smaller proportions at 0.56, 0.29, and 0.13%, respectively.

Shifting our focus to Sindh, 'Protected Wells' continue to play pivotal roles, with a share of 50.56%. 'Piped Water' sources are also substantial, contributing to 36.90% of the population's drinking water. 'Bottled/filtered Water' sources make a 5.89% contribution, while 'Unprotected Wells', 'Surface Water', and 'Commercial Entities' constitute smaller segments at 1.72, 2.60, and 2.33%, respectively.

In Khyber Pakhtunkhwa (KP), 'Piped Water' is the dominant source, providing drinking water to 47.34% of the region's population. 'Protected Wells' closely follow at 35.16%. 'Surface Water' plays a notable role with a 15.17% contribution, while 'Bottled/filtered Water', 'Unprotected Wells', and 'Commercial Entities' are relatively less significant at 0.70, 1.18, and 0.45%, respectively.

Balochistan's water source distribution highlights the significance of 'Piped Water', accounting for 47.13% of the total drinking water. 'Protected Wells' contribute 28.13%, while 'Surface Water' makes up 11.78%. 'Bottled/filtered Water', 'Unprotected Wells', and 'Commercial Entities' hold smaller portions at 1.26, 4.08, and 7.62%, respectively.

In Gilgit-Baltistan (GB), 'Piped Water' takes center stage as the primary source, with a remarkable 67.67% share. 'Protected Wells' play minor roles at 6.10%, while 'Surface Water' is relatively significant at 23.88%. 'Unprotected Wells', 'Bottled/filtered Water', and 'Commercial Entities' make smaller contributions at 0.19, 1.33, and 0.83%, respectively.

The ICT showcases 'Protected Wells' as the primary source, contributing 58.57% of the drinking water. 'Bottled/filtered Water' sources are substantial, accounting for 14.26%. 'Piped Water' contributes 18.70%, while 'Surface Water', 'Unprotected Wells', and 'Commercial Entities' have smaller shares at 0.77, 2.83, and 4.87%, respectively.

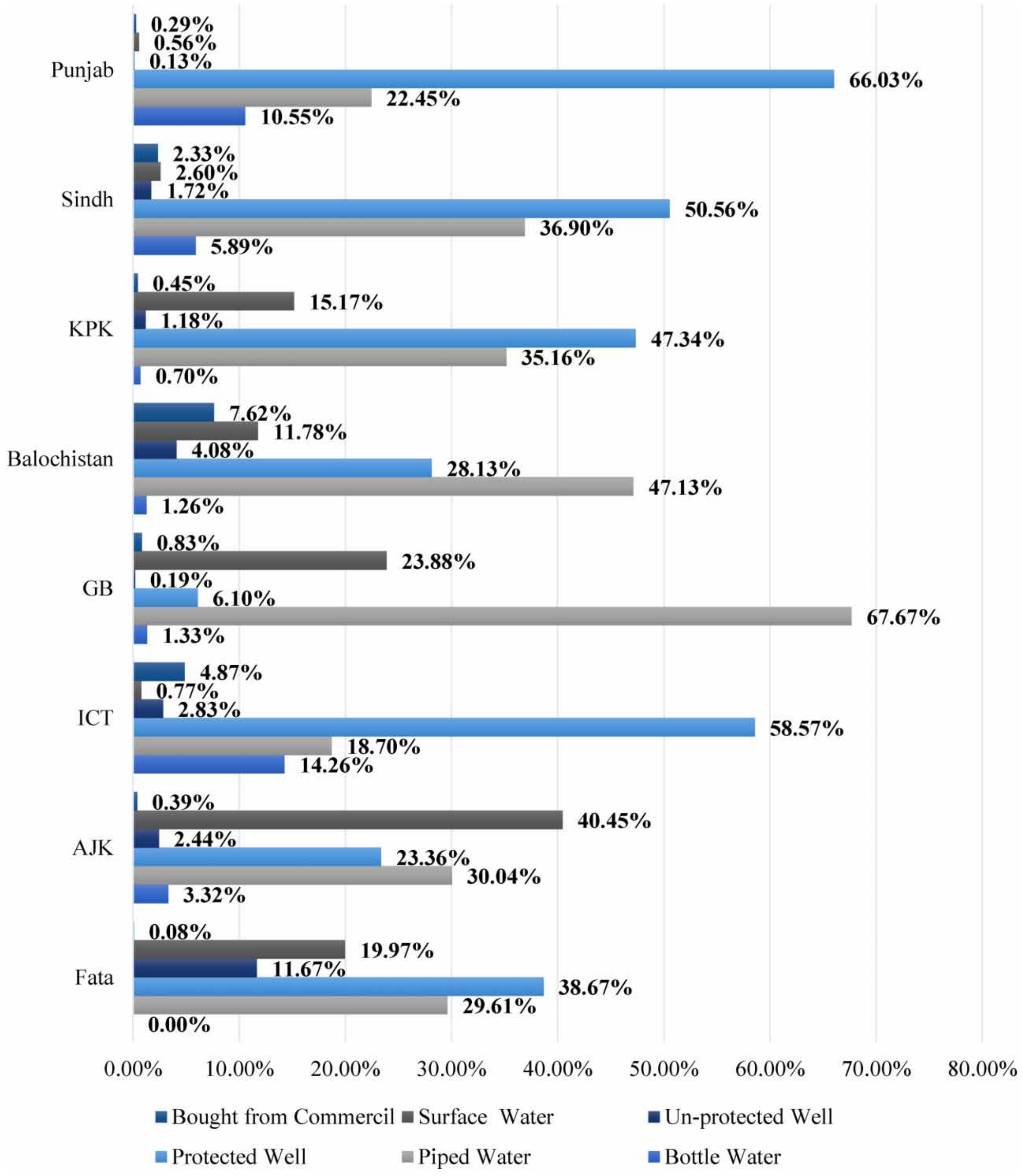


Figure 4 | Weighted percentages of drinking water sources.

In AJK, 'Surface Water' takes the lead as the dominant source, providing 40.45% of the drinking water. 'Piped Water' is also substantial at 30.04%. 'Protected Wells' contribute 23.36%, while 'Unprotected Wells', 'Bottled/filtered Water', and 'Commercial Entities' have smaller percentages at 2.44, 3.32, and 0.39%, respectively.

Finally, Federally Administered Tribal Areas (FATA) primarily relies on 'Protected Wells', making up 38.67% of the drinking water. 'Piped Water' contributes 29.61%. 'Surface Water' is notable at 19.97%, while 'Unprotected Wells', 'Bottled/filtered Water', and 'Commercial Entities' play smaller roles, with 'Commercial Entities' being the least utilized at 0.08%.

Figure 5 shows the percentage of people in different regions of Pakistan who practice water purification. The data are divided into two categories: YES (those who purify their water) and NO (those who do not).

The highest rate of water purification is in AJK (36.05%), followed by Punjab (16.29%) and Sindh (21.21%). The lowest rate of water purification is in the ICT (2.84%), followed by Balochistan (5.75%) and KP (6.74%).

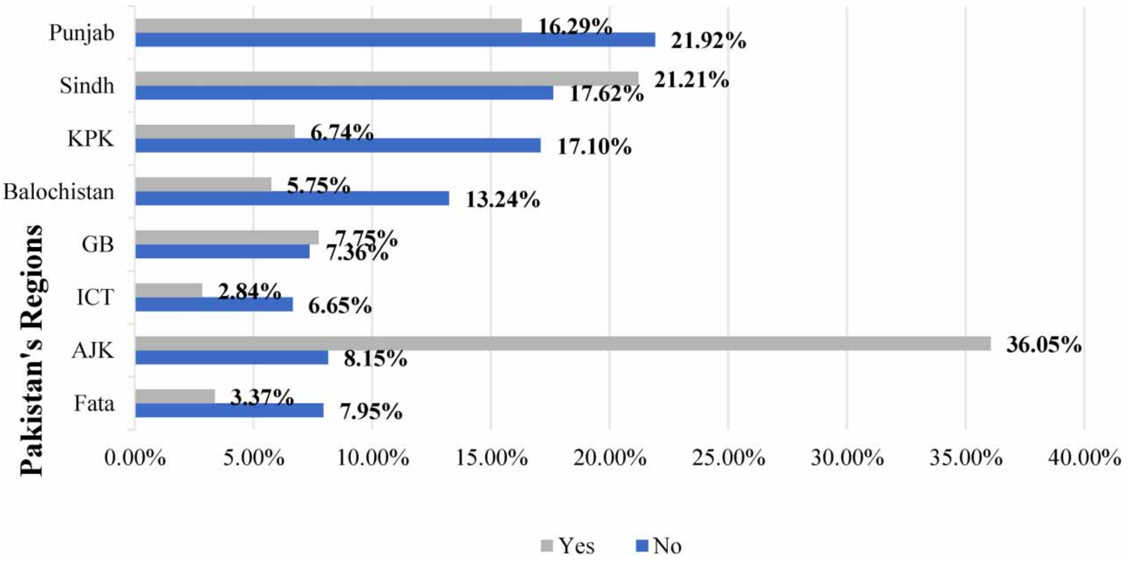


Figure 5 | Unweighted percentage of water purification with respect to region.

These findings suggest that there are significant regional differences in water treatment behaviors in Pakistan. This could be due to a number of factors, such as access to clean water, awareness of the importance of water purification, and cultural practices.

The data also highlight potential areas for intervention and awareness campaigns. For example, in regions with a low rate of water purification, there is a need to educate people about the importance of purifying their water and to provide them with access to affordable and effective water purification methods.

Figure 6 shows the weighted percentages of people who practice water purification in different regions of Pakistan. Weighted percentages take into account the population size of each region, so they provide a more accurate picture of the overall prevalence of water purification practices in the country. The data show that Punjab has the highest rate of water purification practice (50.39%), followed by Sindh (35.73%), KP (8.05%), Balochistan (4.26%), the ICT (0.38%), and the FATA (1.18%). These findings suggest that there is a significant variation in water purification practices across different regions of Pakistan. This could be due to several factors, such as access to clean water, awareness of the importance of water purification, and cultural practices.

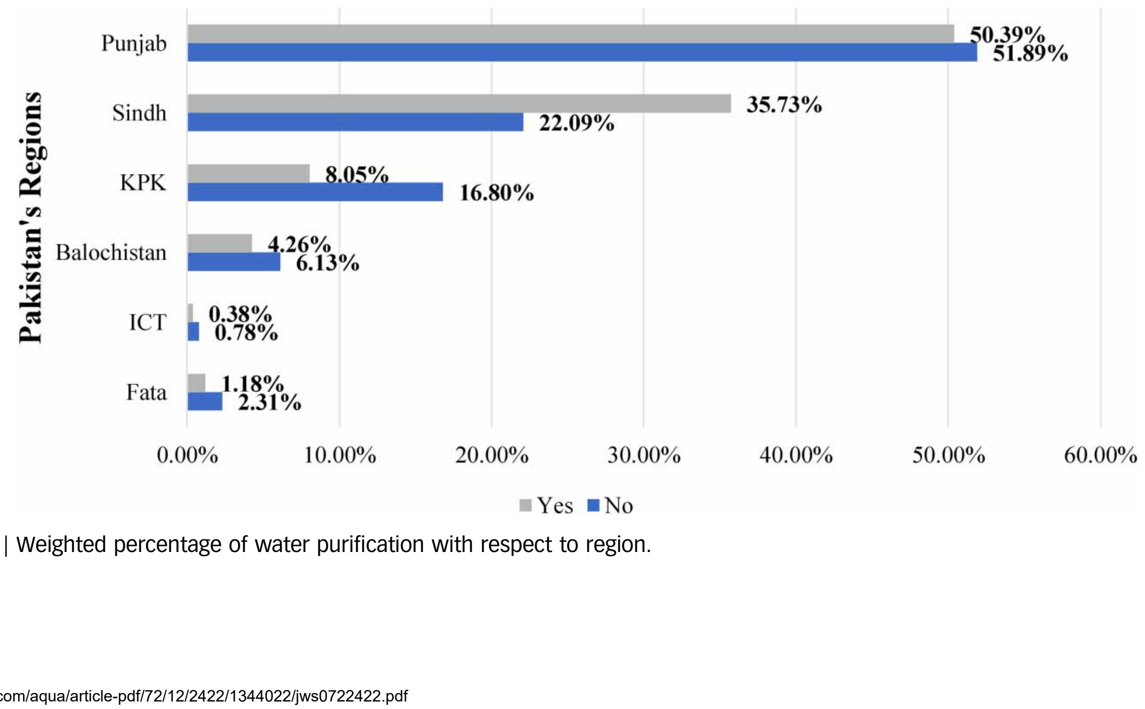


Figure 6 | Weighted percentage of water purification with respect to region.

It is also worth noting that the overall rate of water purification practice in Pakistan is relatively low, with only 36% of the population reporting that they purify their water. This suggests that there is a need to promote the importance of water purification and to make affordable and effective water purification methods more accessible to the population

3.2. Practical implications

This study is crucial for public health and water management authorities. It emphasizes the need for customized approaches that consider regional differences and people's economic situations. With the help of this study, policymakers can improve how they allocate resources by focusing on areas with more unsafe water sources, making their efforts more effective. The study also shows that education and exposure to media have a big impact on safe water drinking habits, this means that it is important to educate people in areas where education levels are lower. Getting communities involved, especially in places where safe water drinking habits are less common, is very important. When communities work together, it can encourage people to use safer water sources and treatments. The study also shows that when water quality is better, people are healthier. So, it is important to make sure healthcare resources are used wisely, especially in places where unsafe water is a bigger problem. Because of climate change, we need to think about how it affects our water sources. This research helps us plan for stronger water systems that can handle changes in the weather. It can also help us improve our current policies on safe drinking water. Policymakers can look at their policies and change them based on what this study found. By doing these things, this study can help more people in Pakistan have better access to safe drinking water, which will make them healthier.

This study was conducted to identify the factors that influence people's choices of drinking water sources and HWT practices in Pakistan. The study used a mixed-methods approach, including descriptive statistics, inferential testing, and visualizations, to analyze both unweighted and weighted data. The study's findings revealed significant variations in water source preferences and water treatment behaviors across different regions and socioeconomic groups. For example, residents of rural areas were more likely to rely on unimproved water sources, while those with higher incomes were more likely to use improved water sources and practice HWT. The study concludes that the findings can inform targeted interventions and policies aimed at improving water quality and sanitation practices in Pakistan. For example, the findings suggest that there is a need to increase awareness of the importance of using improved water sources and practicing HWT, particularly in rural areas and among low-income groups. The findings also suggest that the government should invest in improving access to improved water sources and provide affordable water treatment products and services. In simpler terms, the study found that people in rural areas and people with less money are more likely to drink unsafe water. The study recommends that the government and other organizations work to make safe water more accessible to everyone in Pakistan.

4. CONCLUSION AND FUTURE DIRECTION

In developing countries, the compromised quality of potable water stands as a critical health concern due to its association with various severe illnesses, notably diarrhea and hepatitis. This research aimed to examine how distinct socioeconomic factors within households influence the selection of water sources and the adoption of HWT practices, with a particular focus on weighted and unweighted analyses. The study outcomes offer valuable insights for policymakers to address barriers hindering the consumption of safe drinking water in Pakistan. Weighted analysis, which accounts for the complex survey design, reveals nuanced patterns that may be overlooked in unweighted approaches. In this context, the findings show that larger household sizes and increased media exposure are associated with a decreased probability of using bottled or filtered water for safe drinking. Conversely, higher levels of education, women's empowerment in household purchases, and greater household wealth are positively correlated with a preference for bottled or filtered water. Moreover, the analysis demonstrated that rural households, older household heads, individuals with large family sizes, lower education levels, and lower wealth statuses are less inclined to adopt HWT methods in the weighted model. However, when considering proximity to a water source belonging to wealthier quintiles, it becomes a significant predictor of higher HWT adoption rates in the weighted context. Significant variations exist between urban and rural populations, as evident in the weighted and unweighted analyses. In the weighted model, urban residents account for 53.05% of the population, while the rural population comprises 45.95%. When considering water sources, in the weighted model, 46.59% predominantly rely on piped water, with only 1.59% opting for bottled/filtered water. HWT adoption stands at 5.69% in the weighted analysis. The 60+ age group constitutes 20.60% of the population in the weighted context. Approximately 65.46% of the population lack formal education, and the poorest segment represents 19.32%. In terms of media exposure, it is observed in 59.48% of the population in the weighted model. The choice of water source also exhibits regional variations. In Punjab, protected wells are the most common source (66.03%) in the weighted analysis, while Sindh relies on protected wells (50.57%) and piped water (36.90%). KP predominantly utilizes protected wells (47.34%) and piped water (35.16%). Balochistan's primary sources are piped water (47.13%) and protected wells (28.13%). GB primarily depends on piped water (67.67%), ICT primarily uses protected wells (58.57%), and AAJK utilizes surface water (40.45%) and protected wells (23.36%). FATA rely on piped water (29.61%) and protected wells (38.67%). The findings reveal that piped water and protected wells are the predominant sources of drinking water in Pakistan, and this trend remains consistent across both weighted and unweighted estimates. Surface water and commercially sourced water play a less prominent role in meeting the drinking water needs of households. These results offer valuable insights into the multifaceted dynamics influencing the selection of drinking water sources and the adoption of HWT practices in Pakistan, highlighting the importance of taking into account the weighted and unweighted analyses when formulating interventions and awareness campaigns to promote safer drinking water practices. Moreover, they underscore the significance of factors like education, wealth, media exposure, and women's empowerment in shaping drinking water preferences. The findings from the MNL model underline the critical role of improving access to piped water and protected water sources in encouraging the adoption of bottled or filtered water. Simultaneously, efforts should be directed toward enhancing education, media exposure, women's empowerment, and wealth to reduce reliance on potentially unsafe water sources, reinforcing the necessity of considering both weighted and unweighted analyses in health policy planning.

The study is inherently region-specific, as it delves into the intricate dynamics of safe drinking water access within the context of Pakistan. While regional in focus, it is essential to recognize that the implications of this research transcend local boundaries. Pakistan holds a distinctive position on the global stage as the fifth most populous nation, and it grapples with the imminent challenges of climate change, which are intricately linked to water resource management. The country's vulnerability to water scarcity issues underscores its significance as a pivotal candidate for in-depth water resource research. Additionally, the study's utilization of data from the PDHS should be acknowledged. It is noteworthy that the Demographic and Health Survey (DHS) series is conducted in various countries, providing a wealth of valuable information. However, it is crucial to emphasize that the concept of weights, a cornerstone of this study, has not been widely incorporated into the analysis of DHS data in the past. The introduction of weighted analyses in the context of safe drinking water and HWT practices offers a novel perspective, enhancing the robustness and accuracy of the findings, not only for Pakistan but potentially serving as a precedent for similar research endeavors in other regions. This innovative approach paves the way for future studies, enriching the methodological toolkit for examining water-related challenges and public health concerns.

In the following, we present some of the future directions and recommendations to enhance the scope and impact of our research.

(i) Given the preference for piped water among urban households and the reliance on surface water and unprotected wells in rural areas, policymakers should prioritize investments in water infrastructure. Improving access to safe and reliable piped water in both urban and rural regions could significantly improve water quality and reduce health risks.

(ii) The study reveals that households with higher education levels have better odds of accessing safe water sources. Therefore, there is a need to promote education and awareness campaigns to increase knowledge about the importance of using safe water sources.

(iii) The study shows that households with higher wealth have better odds of accessing safe water sources. Therefore, there is a need to address poverty and inequality and provide financial support to households that cannot afford safe water sources.

(iv) Since media exposure is linked to reduced reliance on all water sources, policymakers should consider launching targeted awareness campaigns to educate communities about the importance of safe drinking water practices. These campaigns could emphasize the health risks associated with untreated water sources and promote the benefits of adopting HWT methods.

(v) The study shows that households relying on unprotected wells have higher odds of using bottled or filtered water. Therefore, there is a need to invest in water supply infrastructure and improve access to protected water sources to reduce reliance on unprotected water sources.

(vi) The research findings indicate that most households in Pakistan obtain their drinking water from wells. However, the excessive use of wells and tube wells has resulted in a significant reduction in groundwater levels. Consequently, it is essential for the government to launch awareness campaigns to promote the use of filtered and piped water for drinking purposes.

(vii) Since rural areas show higher adoption rates of HWT, policymakers should build on this trend by promoting HWT methods. Information campaigns and training programs on effective and affordable HWT techniques can help improve water quality at the household level.

(viii) Collaborating with local organizations and communities can facilitate the implementation of water quality improvement initiatives. Capacity-building efforts and partnerships with non-governmental organizations can help ensure the sustainability and effectiveness of policies.

(ix) For future research, considering more advanced analysis methods, as seen in recent studies, would allow for a valuable comparison with the weighted estimates presented in this paper. Such comparisons could enhance the comprehensiveness of water resource and quality analysis and lead to more robust conclusions.

AUTHORS CONTRIBUTIONS

All authors have equally contributed to this paper.

DATA AVAILABILITY STATEMENT

All relevant data are available from https://dhsprogram.com/Data/.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Abdalrahman, G., Lai, S. H., Kumar, P., Ahmed, A. N., Sherif, M., Sefelnasr, A., Chau, K.-W. & Elshafie, A. 2022 Modeling the infiltration rate of wastewater infiltration basins considering water quality parameters using different artificial neural network techniques. *Engineering Applications of Computational Fluid Mechanics* **16** (1), 397–421.

Ahmad, I. & Sattar, A. 2010 Factors determining public demand for safe drinking water (A Case Study of District Peshawar).

Ahmad, M., Iftikhar, M. N., Shakeel, K. & Cheema, K. H. 2019 Governance and civic capacity for the provision of drinking water in urban Sindh. US-Pakistan Center for Advanced Studies in Water (USPCAS-W), MUET, Jamshoro, Pakistan All rights reserved by USPCAS-W. The authors encourage fair use of this material for non-commercial purposes with proper citation.

Ahmed, A. & Shafique, I. 2019 Perception of household in regards to water pollution: An empirical evidence from Pakistan. *Environmental Science and Pollution Research* **1** (26), 8543–8551.

Akbar, A., Sitara, U., Khan, S. A., Muhammad, N., Khan, M. I., Khan, Y. H. & Kakar, S. U. 2013 Drinking water quality and risk of waterborne diseases in the rural mountainous area of Azad Kashmir Pakistan. *International Journal of Biosciences* **3** (12), 245–251.

Alvisi, S., Mascellani, G., Franchini, M. & Bardossy, A. 2006 Water level forecasting through fuzzy logic and artificial neural network approaches. *Hydrology and Earth System Sciences* **10** (1), 1–17.

Bibi, T., Khan, H. & Khan, M. A. 2014 Estimation of demand for bottled water in Peshawar–Pakistan. *Euro-Asian Journal of Economics and Finance* **2** (3), 216–223.

Daud, M. K., Nafees, M., Ali, S., Rizwan, M., Bajwa, R. A., Shakoor, M. B., Arshad, M. U., Chatha, S. A., Deeba, F., Murad, W. & Malook, I. 2017 Drinking water quality status and contamination in Pakistan. *BioMed Research International* **5**, 7908183.

Dinka, M. O. 2018 Safe drinking water: concepts, benefits, principles, and standards. *Water Challenges of an Urbanizing World* **163**.

du Preez, M., Conroy, R. M., Ligondo, S., Hennessy, J., Elmore-Meegan, M., Soita, A. & McGuigan, K. G. 2011 Randomized intervention study of solar disinfection of drinking water in the prevention of dysentery in Kenyan children aged under 5 years. *Environmental Science & Technology* **45** (21), 9315–9323.

Ebtehaj, I., Sammen, S. S., Sidek, L. M., Malik, A., Sihag, P., Al-Janabi, A. M. S., Chau, K.-W. & Bonakdari, H. 2021 Prediction of daily water level using new hybridized GS-GMDH and ANFIS-FCM models. *Engineering Applications of Computational Fluid Mechanics* **15** (1), 1343–1361.

Fadhullah, W., Imran, N. I., Ismail, S. N., Jaafar, M. H. & Abdullah, H. 2022 Household solid waste management practices and perceptions among residents in the East Coast of Malaysia. *BMC Public Health* **22** (1), 1–20.

Haydar, S., Arshad, M. & Aziz, J. A. 2009 Evaluation of drinking water quality in urban areas of Pakistan: A case study of Southern Lahore. *Pakistan Journal of Engineering and Applied* **5**, 16–23.

Jalan, J., Somanathan, E. & Chaudhuri, S. 2009 Awareness and the demand for environmental quality: survey evidence on drinking water in urban India. *Environment and Development Economics* **14** (6), 665–692.

Janmaat, J. 2007 A little knowledge…: Household water quality investment in the Annapolis valley. *Canadian Journal of Agricultural Economics/Revue Canadienne D'agroeconomie* **55** (2), 233–253.

Kausar, S., Asghar, K., Anwar, S. M., Shaukat, F. & Kausar, R. 2011 Factors affecting drinking water quality and human health at household level in Punjab, Pakistan. *Statistics* **100** (100), 0.

Lantagne, D. S., Quick, R. & Mintz, E. D. 2006 Household water treatment and safe: Storage options in developing countries. *Navig* **99**, 17–38.

Li, P., Li, G., Guo, L. & Shi, D. 2021 A study on sediment avoidance diversion and the coordinated dispatch of water and sediment at an injection-water supply project on a sediment-laden river. *Engineering Applications of Computational Fluid Mechanics* **15** (1), 530–548.

Luby, S. E., Syed, A. H., Atiullah, N., Faizan, M. K. & Fisher-Hoch, S. 2000 Limited effectiveness of home drinking water purification efforts in Karachi, Pakistan. *International Journal of Infectious Diseases* **4** (1), 3–7.

Memon, M., Soomro, M. S., Akhtar, M. S. & Memon, K. S. 2011 Drinking water quality assessment in Southern Sindh (Pakistan). *Environmental Monitoring and Assessment* **177**, 39–50.

Morakinyo, O. M., Adebowale, S. A. & Oloruntoba, E. O. 2015 Wealth status and sex differential of household head: implication for source of drinking water in Nigeria. *Archives of Public Health* **73** (1), 1–9.

Murtaza, G. & Zia, M. H. 2012 Wastewater production, treatment and use in Pakistan. In: *Second Regional Workshop of the Project 'Safe use of Wastewater in Agriculture*. University of Agriculture, Faisalabad, Pakistan, pp. 16–18.

Pickering, A. J., Crider, Y., Sultana, S., Swarthout, J., Goddard, F. G., Islam, S. A., Sen, S., Ayyagari, R. & Luby, S. P. 2019 Effect of in-line drinking water chlorination at the point of collection on child diarrhoea in urban Bangladesh: A double-blind, cluster-randomised controlled trial. *The Lancet Global Health* **7** (9), e1247–e1256.

Wedgworth, J. C., Brown, J., Johnson, P., Olson, J. B., Elliott, M., Forehand, R. & Stauber, C. E. 2014 Associations between perceptions of drinking water service delivery and measured drinking water quality in rural Alabama. *International Journal of Environmental Research and Public Health* **11** (7), 7376–7392.

Wilson, A. & Boehland, J. 2005 Small is beautiful US house size, resource use, and the environment. *Journal of Industrial Ecology* **9** (1-2), 277–287.

Winship, C. & Radbill, L. 1994 Sampling weights and regression analysis. *Sociological Methods & Research* **23** (2), 230–257.

Wood, S., Foster, J. & Kols, A. 2012 Understanding why women adopt and sustain home water treatment: Insights from the Malawi antenatal care program. *Social Science & Medicine* **75** (4), 634–642.

Zabo, S., Hossain, M. S., Adger, W. N., Matthews, Z., Ahmed, S., Lázár, A. N. & Ahmad, S. 2016 Soil salinity, household wealth and food insecurity in tropical deltas: Evidence from south-west coast of Bangladesh. *Sustainability Science* **11**, 411–421.

Zahid, J. 2018 Impact of clean drinking water and sanitation on water borne diseases in Pakistan.

Zhu, S., Hrnjica, B., Ptak, M., Choiński, A. & Sivakumar, B. 2020 Forecasting of water level in multiple temperate lakes using machine learning models. *Journal of Hydrology* **585**, 124819.

Zoungrana, T. D. 2021 The effect of wealth on the choice of household drinking water sources in West Africa. *International Journal of Finance & Economics* **26** (2), 2241–2250.

Zulfiqar, H., Abbas, Q., Raza, A. & Ali, A. 2016 Determinants of safe drinking water in Pakistan: A case study of faisalabad. *Journal of Global Innovations in Agricultural and Social Sciences* **4** (01), 40–45.

First received 1 September 2023; accepted in revised form 20 November 2023. Available online 30 November 2023